1) Naive Bayes Classifier for Text-base Movie Review Classification

MNBC accuracy = 0.81
MultinomialNB accuracy = 0.81
BNBC accuracy = 0.4066666666666667

ThetaPos, ThetaNeg, ThetaPosTrue and ThetaNegTrue were all attached as separate text files as they were simply too long to fit into the write up.

2) Sample Exam Questions

Sample Exam Questions:

1) For parts (a) and (b), assume we are using a naive Bayes classifier to predict the value of G from the values of the other variables.
a) According to the naive Bayes classifier, what is P(G = 1|a = 1 ∧ b = 1)?

Answer:
P(G = 1 | a = 1 ^ b = 1)
=P(a = 1 ^ b = 1 | G = 1)*P(G=1)/P(a = 1 ^ b = 1)
= 1/8 / 1/4
= 1/2

b) (True/False) Naive Bayes Classifier and logistic regression both directly model p(C|X).

Answer: False, logistic regression is discriminative so it directly models p(C|X), but naive Bayes classifier is generative and first has to make a model of the joint probabilities and doesn't model p(C|X) directly.

c) (True/False) Gaussian Naive Bayes Classifier and Gaussian Mixture Model are similar since both assume that p(X|cluster == i) follows Gaussian distribution.

Answer: True, because the naive Bayes classifier assumes Gaussian distribution and even though GMM uses a mixture of Gaussian distributions, it still assumes a Gaussian distribution for a particular cluster.

d) Answer: False, because with logistic regression it would be linear, but we would get a non-linear boundary for the data shown below using Gaussian naive Bayes classifier.

2)

a) How would a naive Bayes classifier predict y given this input: A = 0, B = 0, C = 1. Assume that in case of a tie the classifier always prefers to predict 0 for y.

Answer: P(y|A=0 ^ B=0 ^ C=1)
= P(y)*(P(A=0|y)*P(B=0|y)*P(C=1|y)/P(A=0 ^ B=0 ^ C=1)

P(y=0|A=0 ^ B=0 ^ C=1)
= 3/7*2/3*1/3*1/3/1/7
= 2/9

P(y=1|A=0 ^ B=0 ^ C=1)
= 4/7*1/4*1/2*1/2/1/7
= 1/4 > 2/9

Since P(y=1|A=0 ^ B=0 ^ C=1) > P(y=0|A=0 ^ B=0 ^ C=1), the classifier will predict 1 for y.

b) Yes, it is possible since even though A, B, and C are independent, they may be dependent in the context of a certain class and therefore other classifiers may be able to do better.

3)

a) Not enough info, need P(A).
b) Not enough info, need P(A).
c) Not enough info, need P(A).
d) Yes, there is enough info because with Bayes' rule we know that P(B|A) = P(B)*P(A|B)/P(A). So to calculate it we do 1/3 * 2/3 / 4/9 = 1/2.